

Calculation of the scalar diffraction field from curved surfaces by decomposing the three-dimensional field into a sum of Gaussian beams

Erdem Şahin* and Levent Onural

Department of Electrical and Electronics Engineering, Bilkent University, TR-06800 Bilkent, Ankara, Turkey

**Corresponding author: sahin@ee.bilkent.edu.tr*

Received December 19, 2012; revised January 28, 2013; accepted February 1, 2013;
posted February 4, 2013 (Doc. ID 181970); published February 28, 2013

We present a local Gaussian beam decomposition method for calculating the scalar diffraction field due to a two-dimensional field specified on a curved surface. We write the three-dimensional field as a sum of Gaussian beams that propagate toward different directions and whose waist positions are taken at discrete points on the curved surface. The discrete positions of the beam waists are obtained by sampling the curved surface such that transversal components of the positions form a regular grid. The modulated Gaussian window functions corresponding to Gaussian beams are placed on the transversal planes that pass through the discrete beam-waist position. The coefficients of the Gaussian beams are found by solving the linear system of equations where the columns of the system matrix represent the field patterns that the Gaussian beams produce on the given curved surface. As a result of using local beams in the expansion, we end up with sparse system matrices. The sparsity of the system matrices provides important advantages in terms of computational complexity and memory allocation while solving the system of linear equations. © 2013 Optical Society of America

OCIS codes: 070.7345, 090.0090, 090.1760, 090.1995.

1. INTRODUCTION

The commonly used source model approaches [1–7] do not produce accurate results for the problem of three-dimensional (3D) field calculation from the two-dimensional (2D) field specified on a curved surface (or object). The reason is that they ignore the mutual couplings between the field samples on the given curved surface. As a result, the subsequently reconstructed field on the same curved surface is found to be inconsistent with the original one [8,9].

The local Gaussian beam decomposition proposed in [8] reduces the mutual coupling effects with respect to point source models [1–7]. In that work of Şahin and Onural [8], the signal given on the curved surface is decomposed into a sum of modulated Gaussian window functions. Then the 3D field is written as a sum of 3D Gaussian beams, each of which corresponds to a modulated Gaussian window function on the curved surface. Although the method proposed by [8] reduces the mutual couplings compared to point-source models considerably, for some surfaces this method still produces inconsistent 3D field solutions. The reason is that the disjoint patches that represent the supports of the modulated Gaussian window functions on the surface are treated independently. Thus, possible mutual couplings between such patches are omitted. As a result of that, the method proposed in [8] is applicable to the surfaces such that the mutual couplings between the disjoint patches on the surface are negligible. In addition to this, a smoothness constraint is imposed on the curved surface for the applicability of the formulations developed in [8]. The mutual coupling problem is solved by using a field model approach, based on the plane-wave decomposition (PWD), in [9,10], resulting in an exact method to calculate the diffraction field from a given curved surface. However, for

large sizes of curved surfaces, this method becomes impractical because of an intolerable increase in the computational complexity and the related memory requirement.

Here in this paper, we reformulate the local Gaussian beam decomposition method introduced in [8] in order to achieve the exact field solutions. The 3D Gaussian beam decomposition method that we propose does not impose any restriction on the curved surface and still produces consistent 3D field solutions. We write the 3D field as a sum of 3D beams whose waist positions are determined by the given curved surface. The directions of the beams are fixed independent of the surface. We find the coefficients of the beams by constructing the linear system of equations where the system matrix is formed by calculating the field patterns that the Gaussian beams produce on the curved surface. In this sense, the formulation of the proposed method is similar to the field model approach given by [9]. However, the locality of the signals used in the proposed decomposition results in important advantages over the approaches that use global signal decomposition methods, for instance [9], in terms of computational complexity and memory allocation.

In Section 2, we start by giving the Gaussian beam decomposition for planar input surfaces. Then we introduce the proposed Gaussian beam decomposition for curved input surfaces in Section 3. We test our approach and present the simulation results in Section 4. Finally, we give the conclusions in Section 5.

2. DIFFRACTION FIELD CALCULATION FROM PLANAR INPUT SURFACES

In this section, we will give the 3D field expression corresponding to the field specified on a planar surface by using

the Gaussian beam decomposition method. Before doing this, let us define the signal space of the monochromatic scalar 3D fields that we consider in this paper as [11]

$$u(x) = \iint_B A(f_x, f_y) \exp\{j2\pi(f_x^T x)\} df_x df_y, \quad (1)$$

where $x = [x, y, z]^T \in \mathbb{R}^3$, $f_x = [f_x, f_y, f_z]^T \in \mathbb{R}^3$, $f_x^2 + f_y^2 + f_z^2 = (1/\lambda^2)$, and λ is the wavelength of the monochromatic light. Note that Eq. (1) is actually called as plane wave decomposition, where the 3D field is written as a sum of infinite-extent plane waves that propagate toward different directions. Here, we take into account only the propagating plane waves whose frequency components along the x and y axes satisfy $f_x^2 + f_y^2 \leq (1/\lambda^2)$. In other words, we exclude the evanescent waves. The propagating plane waves occupy the 2D circular spatial frequency band that is centered at the origin and has a radius $1/\lambda$; we call this band B . In this paper, a given 3D field $u(x): \mathbb{R}^3 \rightarrow \mathbb{C}$ is assumed to be formed by a superposition of such propagating plane waves with $f_z > 0$ [12]. In other words, the z component of the propagation direction is always positive. The coefficient of each propagating plane wave can be found by using the field $u_0(x, y) = u(x, y, 0)$ given on the reference input plane at $z = 0$ as

$$A(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) \exp\{-j2\pi(f_x x + f_y y)\} dx dy, \quad (2)$$

where the 2D function $u_0(x, y)$ is a band-limited function such that its spectrum is zero outside the circular frequency band defined by $f_x^2 + f_y^2 \leq (1/\lambda^2)$. The coefficients given by Eq. (2) actually correspond to the Fourier transform of the 2D input field $u_0(x, y)$ given on the $z = 0$ plane, where the Fourier transform, from the (x, y) domain to the (f_x, f_y) domain, of a 2D function $p(x, y)$ is defined as

$$P(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \exp\{-j2\pi(f_x x + f_y y)\} dx dy. \quad (3)$$

In order to write the 3D field as a sum of Gaussian beams, let us first write the corresponding Gaussian signal decomposition on the input plane. In the continuous case, the shift positions and modulation frequencies of the Gaussian window functions are continuously parameterized. The continuously parameterized Gaussian window function decomposition of the field $u_0(x, y)$ specified on the planar surface at $z = 0$ is written in the form of a continuous short time Fourier decomposition as [13]

$$u_0(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\xi, \eta, f_x, f_y) g(x - \xi, y - \eta) \times \exp\{j2\pi(f_x x + f_y y)\} d\xi d\eta df_x df_y, \quad (4)$$

where $g(x, y) = ce^{-(x^2 + y^2)/\sigma^2}$ is a unit-energy Gaussian window function. Note that c is a constant, which makes $g(x, y)$ a unit energy function. A possible way to find the coefficients of the Gaussian window functions is given by

$$a(\xi, \eta, f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) g^*(x - \xi, y - \eta) \times \exp\{-j2\pi(f_x x + f_y y)\} dx dy. \quad (5)$$

Defining the frequency support of a Gaussian window function as the region outside of which the magnitude of its spectrum is effectively zero, we can say that the coefficients of the Gaussian window functions whose frequency supports are outside the region defined by $f_x^2 + f_y^2 \leq (1/\lambda^2)$ are effectively zero. This is a consequence of the band-limitedness of $u_0(x, y)$.

Note that in Eq. (4) the input signal $u_0(x, y)$ specified on the $z = 0$ plane is decomposed into a sum of infinitely many modulated and shifted versions of the Gaussian window function $g(x, y)$. With a proper discretization of positions and modulation frequencies, the discrete parameter version of Eq. (4) can be written by taking into account only a finite number of modulated and shifted versions of $g(x, y)$ as [14]

$$u_0(x, y) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g(x - mX, y - nY) \times \exp\{j2\pi(kF_x x + lF_y y)\}, \quad (6)$$

where X, Y are the shift steps in space and F_x, F_y are the shift steps in frequency. In the discrete case, an analysis window function $w(x, y)$, corresponding to the Gaussian synthesis window function $g(x, y)$, can be found such that the signal $u_0(x, y)$ is exactly obtained with the coefficients a_{mnkl} [15,16]. Thus,

$$a_{mnkl} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x, y) w^*(x - mX, y - nY) \times \exp\{-j2\pi(kF_x x + lF_y y)\} dx dy. \quad (7)$$

Here note that, for the reasons explained before, the coefficients of some Gaussian window functions are already effectively zero. In [14], the decomposition is restricted to the case that the space-frequency domain is critically sampled ($XF_x = 1, YF_y = 1$). However, for the choice of Gaussian-shaped signal as the synthesis window function, it is shown in [17] that oversampling of the space-frequency domain ($XF_x < 1, YF_y < 1$) should be preferred, if a numerically stable reconstruction is desired.

The 3D field at an observation point $x = [x, y, z]^T$ due to the field $u_0(x, y)$ specified on $z = 0$ plane can be found by summing up the contributions of the Gaussian beams that correspond to the modulated Gaussian window functions on the input plane. Thus,

$$u(x) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(x - p_{mn}). \quad (8)$$

Here, $g_{mnkl}(x - p_{mn})$ is defined as the 3D field expression of the Gaussian beam at the observation point $x = [x, y, z]^T$. Note that the Gaussian beam defined as $g_{mnkl}(x - p_{mn})$ corresponds to the modulated Gaussian window function, which is placed at $p_{mn} = [mX, nY, 0]^T$ on the $z = 0$ plane and has the modulation frequencies kF_x and lF_y along the x and y axes, respectively.

3. DIFFRACTION FIELD CALCULATION FROM CURVED INPUT SURFACES

In Section 2, we developed the 3D field expression corresponding to the 2D field specified on a planar surface. It is clearly seen from Eq. (8) that the 3D field is written as a sum of Gaussian beams that propagate toward different

directions and whose waist positions are placed at different positions on the input plane. In this section, we generalize this approach to the curved surfaces.

Let us consider a curved surface $S \subset \mathbb{R}^3$. We represent the points on S by the vector \mathbf{r} , that is, $\mathbf{r} \in S$. Remember from Section 2 that the 3D field $u(\mathbf{x})$ is uniquely determined by the 2D field defined on the infinite-extent reference plane at $z = 0$, provided that the propagation direction components of the waves along the z axis are always positive. The 2D fields over the planes that are parallel to the reference plane can be found by using the Rayleigh–Sommerfeld diffraction model. As described in [8], what we call as the 3D field is the concatenation of such 2D fields over the planes that are parallel to the reference plane at different depths. The 3D field $u(\mathbf{x})$ must be interpreted as such a 3D field, and $u(\mathbf{r})$ is a 2D field resulting from the intersection of such a 3D field by the curved surface S .

It is possible to find the 3D field due to a field $u(\mathbf{r})$ given on the curved surface S by using Eq. (1) or Eq. (8). In order to do this, one needs to construct the system of linear equations

$$u(\mathbf{r}) = \iint_B A(f_x, f_y) \exp\{j2\pi \mathbf{f}_x^T \mathbf{r}\} d\mathbf{f}_x d\mathbf{f}_y, \quad (9)$$

$$u(\mathbf{r}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{r} - \mathbf{p}_{mn}), \quad (10)$$

corresponding to Eqs. (1) and (8), respectively. Given the field $u(\mathbf{r})$ on S , the coefficients of the expansion elements can be found by solving these equations. Remember from Section 2 that the 3D field $u(\mathbf{x})$ is assumed to be formed by a superposition of propagating plane waves that occupy the frequency band defined by $f_x^2 + f_y^2 \leq (1/\lambda^2)$. Therefore, the coefficients of the Gaussian window functions whose frequency supports are outside the region defined by $f_x^2 + f_y^2 \leq (1/\lambda^2)$ are effectively zero. Thus, such Gaussian window functions need not be taken into account in Eq. (10). Once the coefficients of the expansion functions given by Eqs. (9) and (10) are found, the 3D field at an arbitrary point \mathbf{x} can be calculated by using Eq. (1) or Eq. (8), respectively.

The sparsity of the system matrix related to Eq. (10) is one of the main factors affecting the computational complexity in the solution of Eq. (10). The main benefit of the sparse matrices is that the computational complexity related to the linear system of equations solver is reduced if the solver can take advantage of the percentage and distribution of the zero elements of the system matrix [18,19]. In addition to this, the memory requirement, related to the storage of the system matrix, can be reduced during the solution of the linear system of equations [18–20]. In Eqs. (8) and (10), the waist positions of the Gaussian beams are all chosen on a reference plane. Such an approach is reasonable if this plane is the input surface, as in the case discussed in Section 2, because the width of a Gaussian beam is minimum at the position of beam waist. However, because the Gaussian beams spread (their widths increase) as they propagate beyond the waist position, using the decomposition given by Eq. (8) is not the best choice for curved input surfaces. In this paper, as a more reasonable approach, we choose the waist positions of the Gaussian beams on the curved surface (see Fig. 1). By doing so, the patterns resulting from the intersection of Gaussian beams by the

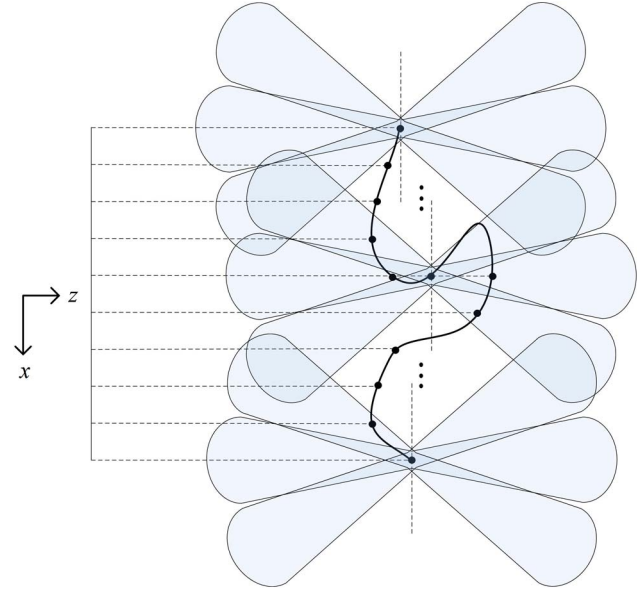


Fig. 1. (Color online) Setup of the proposed Gaussian beam decomposition method (a 2D pattern is shown for the sake of simplicity).

given curved surface become more local than the case where Eq. (8) is used. Hence, noting that the patterns of the Gaussian beams on the curved surface actually represent the columns of the system matrix, more sparse system matrices are obtained. Note that each modulated and shifted version of a Gaussian-shaped elementary signal has the smallest possible support in the space-frequency domain. This is a particularly desirable property for the aim of this paper because it is another factor affecting the sparsity of the system matrix related to Eq. (10).

In the proposed method, we define the modulated Gaussian window functions on the hypothetical planes that are parallel to $z = 0$ plane and pass through the discrete positions $\mathbf{s}_{mn} = [mX, nY, \zeta_{mn}]^T$ on S (see Fig. 1). Here, ζ_{mn} represents the depth of the point on S for the given discrete position $[mX, nY]^T$ on the $z = \zeta_{mn}$ plane. The modulated Gaussian window functions defined on the planes at $z = \zeta_{mn}$ correspond to Gaussian beams in 3D space where \mathbf{s}_{mn} represents the waist positions of these beams. Note that the discrete beam waist positions are obtained by sampling of S using a regular grid on the transversal plane. For some surfaces, after applying such a sampling, there may be multiple discrete points that have the same transversal position but are at different depths (see Fig. 1). For such surfaces, the point that makes the system matrix more sparse is preferable. Using the proposed idea, we find the field at an observation point $\mathbf{x} = [x, y, z]^T$ due to the field $u(\mathbf{r})$ specified on the curved surface S by summing up the contributions of the Gaussian beams as

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{x} - \mathbf{s}_{mn}), \quad (11)$$

where $g_{mnkl}(\mathbf{x} - \mathbf{s}_{mn})$ represents the 3D field expression of the Gaussian beam at the observation point $\mathbf{x} = [x, y, z]^T$ due to the modulated Gaussian window function, which is defined at the point \mathbf{s}_{mn} on the $z = \zeta_{mn}$ plane and has the modulation frequencies kF_x and lF_y along the x and y axes, respectively.

Note that in the proposed method, the transversal waist positions and the discrete set of modulation frequencies are fixed independent of the surface (see Fig. 1). Equation (11) actually defines a 3D signal space for $u(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{C}$. In order that such a signal space covers the signal space defined in Section 2, the discrete shift steps X, Y in the positions and F_x, F_y in the modulation frequencies of the Gaussian window functions should be chosen sufficiently small. The number of Gaussian beams to be included in Eq. (11) depends also on the frequency band $f_x^2 + f_y^2 \leq (1/\lambda^2)$ of the signal space defined in Section 2. Indeed, the Gaussian beams corresponding to Gaussian window functions whose frequency supports are outside the region defined by $f_x^2 + f_y^2 \leq (1/\lambda^2)$ need not be taken into account. Similar to the case where the positions of the Gaussian window functions are taken on a reference plane [see Eq. (10)], we find the coefficients of the Gaussian beams by solving the system of linear equations given as

$$u(\mathbf{r}) = \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{r} - \mathbf{s}_{mn}). \quad (12)$$

We calculate the field at a given point \mathbf{x} by

$$u(\mathbf{x}) = \sum_m \sum_n \sum_k \sum_l \hat{a}_{mnkl} g_{mnkl}(\mathbf{x} - \mathbf{s}_{mn}), \quad (13)$$

where the estimated coefficients \hat{a}_{mnkl} denote the least squares solution of Eq. (12) defined by

$$\hat{a}_{mnkl} = \underset{a_{mnkl}}{\operatorname{argmin}} \left\{ \iint_S |u(\mathbf{r}) - \sum_m \sum_n \sum_k \sum_l a_{mnkl} g_{mnkl}(\mathbf{r} - \mathbf{s}_{mn})|^2 dS \right\}. \quad (14)$$

Here, once more note that by using the proposed Gaussian beam decomposition method explained above, we end up with systems of linear equations, as given by Eq. (12), where the system matrices are sparse. This sparsity is the fact that we take advantage of, and we achieve considerable improvements in terms of computational complexity regarding the solution of the linear system of equations for the problem of scalar diffraction field calculation from curved surfaces. The related improvements are reported in the following section.

4. SIMULATION RESULTS

We show the applicability of the developed formulations for both one-dimensional (1D) curved lines and 2D curved surfaces. Note that the z variable represents the depth in both cases. For the 2D (x, z) -space case, Eqs. (11) and (12) are written as

$$u(\mathbf{x}) = \sum_m \sum_k a_{mk} g_{mk}(\mathbf{x} - \mathbf{s}_m), \quad (15)$$

$$u(\mathbf{r}) = \sum_m \sum_k a_{mk} g_{mk}(\mathbf{r} - \mathbf{s}_m), \quad (16)$$

where $\mathbf{x} = [x, z]^T$ and $\mathbf{s}_m = [mX, \zeta_m]^T$. Note that here the \mathbf{r} vector is used to represent the points on the given curved line.

Assuming that we deal with the fields that are formed by a superposition of propagating plane waves with $f_z > 0$ [see

Eq. (1)], we can uniquely determine such fields given the intersection of these fields by any one of the surfaces having infinite extent along the x and y axes. For the 2D (x, z) -space case, we use the 2D cross sections of such plane waves by the (x, z) plane. Because of numerical concerns, we deal with the fields that are periodic along the x axis, and we intersect such fields by the curved lines that are also periodic with the same period along the x axis. By this way, with a proper discretization, we restrict the signal space so that it is spanned by a finite number of plane waves. We denote the number of plane waves constructing the signal space as N . In order to span such a signal space, we write each Gaussian beam as a sum of appropriately weighted plane waves (which are in the signal space) and obtain the periodic Gaussian beam expression

$$g_{mk}(\mathbf{x} - \mathbf{s}_m) = \frac{1}{N} \sum_{l=-\frac{N}{2}}^{\frac{N}{2}-1} \hat{G}_{mk} \left(\frac{l}{X_p} \right) \times \exp \left\{ j2\pi \left[\frac{l}{X_p} x + \sqrt{\frac{1}{\lambda^2} - \frac{l^2}{X_p^2}} (z - \zeta_m) \right] \right\}, \quad (17)$$

where X_p is the period of the curved line along the x axis. The Fourier transform $\hat{G}_{mk}(f_x)$ of the shifted and modulated version of the Gaussian window function $g(x) = c \exp(x^2/\sigma^2)$ is found as

$$\begin{aligned} \hat{G}_{mk}(f_x) &= \mathcal{F}\{g(x - mX) \exp(j2\pi k F_x x)\} \\ &= \sqrt{c\pi\sigma} \exp\{-\pi^2\sigma^2(f_x - kF_x)^2\} \\ &\quad \times \exp\{-j2\pi(f_x - kF_x)mX\}. \end{aligned} \quad (18)$$

Note that because the spectrums of the Gaussian window functions are local, only a small percentage of plane waves that are in the signal space are sufficient to represent the corresponding Gaussian beams.

We uniformly discretize the given periodic curved line with sampling step L_s and have T samples per period. Denoting the resulting discrete points taken on the curved line as $\mathbf{r}_n = \mathbf{r}(L_s n)$, where $\mathbf{r}(l)$ is the arc-length parameterization of the curved line, we construct the linear system of equations by using Eq. (16) for each discrete point \mathbf{r}_n as

$$u(\mathbf{r}_n) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}-1} \sum_{k=-\frac{K}{2}}^{\frac{K}{2}-1} a_{mk} g_{mk}(\mathbf{r}_n - \mathbf{s}_m), \quad (19)$$

where M is the total number of shift positions in one period and K is the total number of modulation frequencies. Note that the discrete beam-waist positions \mathbf{s}_m are obtained by using a uniform sampling grid on the x axis with a sampling step X . Thus, $\mathbf{s}_m = [mX, \zeta_m]^T$.

Using matrix notation, we write Eq. (19) in a more compact form as

$$\mathbf{U} = \mathbf{G}\mathbf{a}, \quad (20)$$

where \mathbf{U} is the vector that represents one period of the field given on the curved surface, \mathbf{G} is the matrix whose columns represent one period of the field patterns that different parallel Gaussian beams produce on the curved line, and \mathbf{a} is the vector of coefficients of the parallel Gaussian beams to be found. Here, \mathbf{U} and \mathbf{a} are vectors that have lengths T and

N , respectively; \mathbf{G} is a $T \times N$ matrix. Note that the total number of parallel Gaussian beams used in the decomposition is equal to N (the degree of freedom of the signal space), that is, $MK = N$. We form the system matrix \mathbf{G} as

$$\mathbf{G} = [g_{11}(\mathbf{r} - \mathbf{s}_{11}) | g_{12}(\mathbf{r} - \mathbf{s}_{12}) | \dots | g_{MK}(\mathbf{r} - \mathbf{s}_{MK})] \quad (21)$$

and the coefficients vector \mathbf{a} as

$$\mathbf{a} = [a_{11}, a_{12}, \dots, a_{MK}]^T. \quad (22)$$

In order that the exact field can be obtained, the number of samples given on the curved line must be equal to or larger than the number of Gaussian beams used in the decomposition, that is, $T \geq N$. Otherwise, multiple solutions exist. Denoting the least squares solution of the resulting overdetermined system of linear equations, given by Eq. (20), as

$$\hat{\mathbf{a}} = [\hat{a}_{11}, \hat{a}_{12}, \dots, \hat{a}_{MK}]^T \\ = \underset{\mathbf{a}}{\operatorname{argmin}} \left\{ \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \left| u(\mathbf{r}_n) - \sum_m \sum_k a_{mk} g_{mk}(\mathbf{r}_n - \mathbf{s}_m) \right|^2 \right\}, \quad (23)$$

the field at an observation point \mathbf{x} is found as

$$u(\mathbf{x}) = \sum_{m=-\frac{M}{2}}^{\frac{M}{2}-1} \sum_{k=-\frac{K}{2}}^{\frac{K}{2}-1} \hat{a}_{mk} g_{mk}(\mathbf{x} - \mathbf{s}_m). \quad (24)$$

Note that as long as the observation distance z is sufficiently large and the waist of the Gaussian beams is sufficiently small, the Gaussian beam expression introduced in [8] can be used in Eq. (24). The mentioned expression, which is based on the Rayleigh–Sommerfeld diffraction model, gives the scalar 3D field corresponding to a modulated Gaussian window function quite accurately even for nonparaxial cases [8,21].

In order to test the accuracy and efficiency of the developed formulations, we start with a 2D periodic and continuous (x, z) field that is specified by randomly choosing the coefficients of the periodic Gaussian beams from the uniform distribution on the open interval $(0,1)$. Using Eq. (24), we calculate the resulting field on the $z = 0$ line to obtain our test signal. Similarly, we intersect the 2D test field by the curved line at the discrete points \mathbf{r}_n to obtain the \mathbf{U} vector. Using Eq. (17), we find the field samples that each Gaussian beam produces at the discrete points \mathbf{r}_n on the curved line and form the \mathbf{G} matrix. We also construct another overdetermined system of linear equations where the columns of the \mathbf{G} matrix are the field patterns that the plane waves produce on the curved line. We solve these two overdetermined systems of linear equations to find the sets of Gaussian beam and plane-wave coefficients. Using the resulting coefficients, we calculate the samples of the diffraction field on the reference line at $z = 0$ via the Gaussian beam decomposition given by Eq. (24) and PWD given by

$$u(x, z) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \hat{A}_k \exp \left\{ j2\pi \left(\frac{k}{X_p} x + \sqrt{\frac{1}{\lambda^2} - \frac{k^2}{X_p^2}} z \right) \right\}, \quad (25)$$

where $\{\hat{A}_k; k \in \{(-N/2), (-N/2) + 1, \dots, (N/2) - 1\}\}$ is the set of estimated plane-wave coefficients. The accuracies of both methods are verified by comparing the reconstructed signals with the original test signal defined on the $z = 0$ line.

The coefficients of the Gaussian beams and plane waves used in the Gaussian beam decomposition and PWD, respectively, are found via the QR factorization method [22]. In order to test the efficiency of the proposed method, we compare the computation times necessary to solve the two systems of linear equations corresponding to these decompositions. We apply the usual QR factorization routine of MATLAB as well as the suitesparseQR (SPQR) package [18], which provides MATLAB routines to implement multifrontal sparse QR factorization. Note that most of the elements of the \mathbf{G} matrix in the proposed method are practically negligible. Therefore, while using SPQR, we simply apply a thresholding to the \mathbf{G} matrix to obtain a sparse matrix and take advantage of the large number of zero elements of such a sparse matrix.

The simulation setup is seen in Fig. 2. In order to see the effect of the size of the input manifold, we change the period X_p along the x axis and do the simulations for each case. While doing this, we preserve the structure given in Fig. 2. Therefore, as the period of the curved line is increased, the number of samples T taken on the curved line and the number of parallel Gaussian beams N are also increased. We take $T = 1.2N$ and keep the sampling step L_s constant at 0.43λ , where λ is the wavelength of the monochromatic light and taken as 500 nm. Note that by taking $T > N$, we ensure that the resulting systems of linear equations are overdetermined. θ is chosen as 30 deg. The computational complexities of the four operations explained above are given in Fig. 3 with respect to N . The operations mentioned in Fig. 2 are (1) QR factorization is applied to the PWD formulation (QRPWD), (2) multifrontal sparse QR factorization is applied to the PWD formulation (SQRPWD), (3) QR factorization is applied

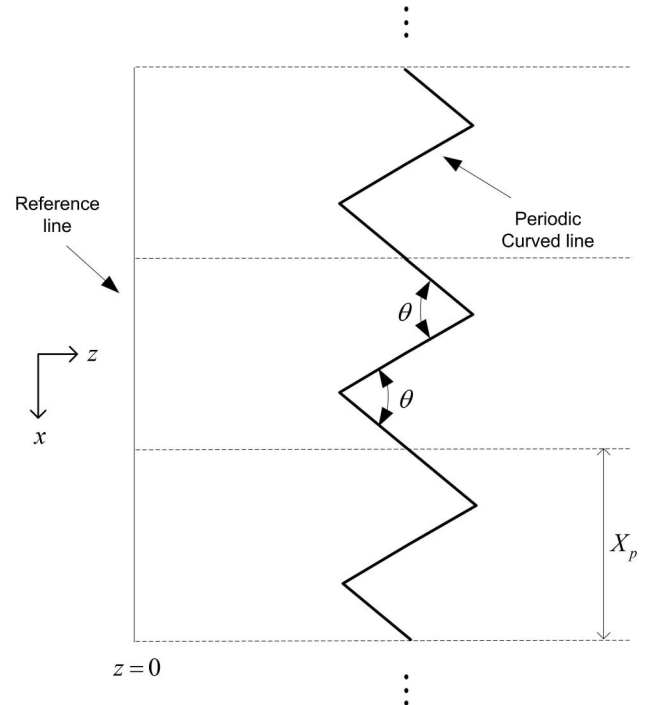


Fig. 2. Periodic 2D simulation setup.

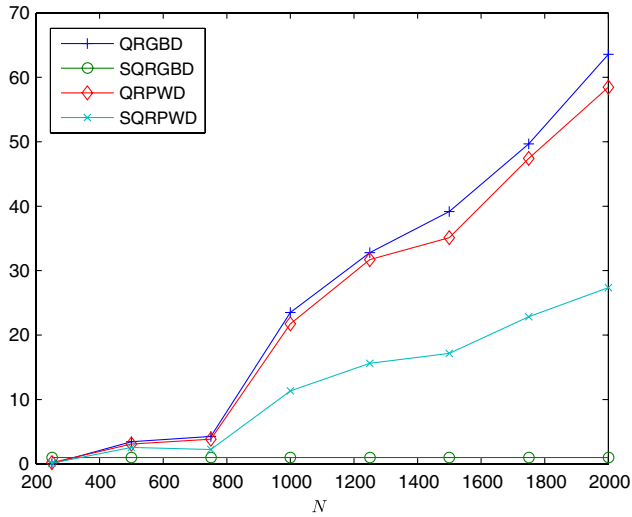


Fig. 3. (Color online) Ratios of the complexities of the operations QRPWD, SQRPWD, and QRGBD to the complexity of SQRGBD (note that the result shown for SQRGBD is “1”).

to the proposed Gaussian beam decomposition formulation (QRGBD), and (4) multifrontal sparse QR factorization is applied to the proposed Gaussian beam decomposition formulation (SQRGBD). Note that the computational complexity of SQRGBD is taken as one unit time as a reference for all T . The computational complexities of other operations represent the relative times with respect to SQRGBD. For the case that the proposed Gaussian beam decomposition method is used, the sparsity of the G matrix is shown in Fig. 4 for the same set of N . Here, the sparsity is defined as the ratio of the number of zero elements of a matrix to its total number of elements. Examples of the system matrices for the proposed Gaussian beam decomposition and the PWD methods are shown in Figs. 5 and 6, respectively, for $T = 2400$. The sparsity of the G matrix corresponding to the proposed method is clearly seen. All these figures justify that as the system matrix corresponding to proposed method gets more sparse, the resulting computational saving becomes more significant. Such a computational advantage makes the proposed method more critical for meaningful sizes of surfaces.

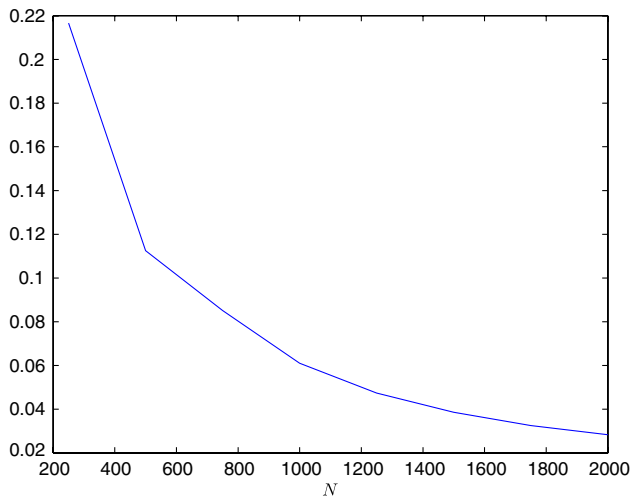


Fig. 4. (Color online) Sparsity of the system matrices for the proposed Gaussian beam decomposition method.

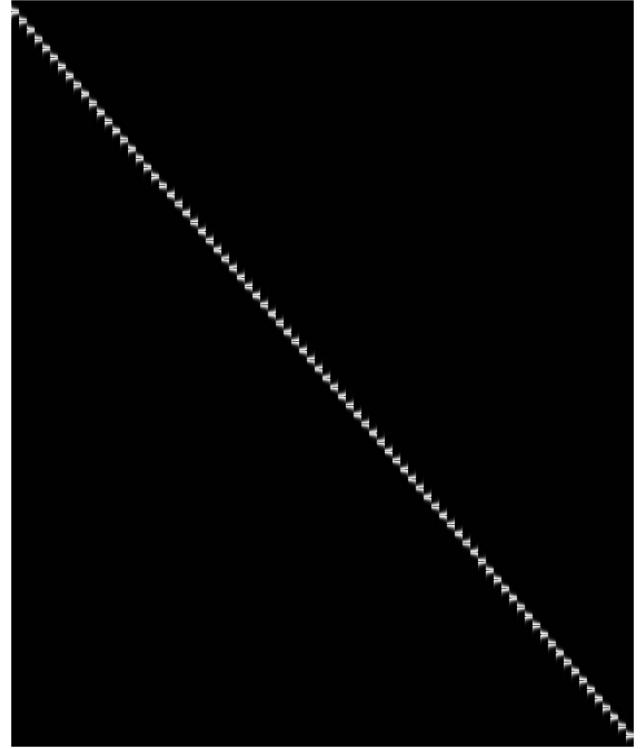


Fig. 5. Absolute value of the system matrix G , which has size 2400×2000 and is formed by the application of the proposed Gaussian beam decomposition method to the curved line shown in Fig. 2.



Fig. 6. Real part of the system matrix G , which has size 2400×2000 and is formed by the application of the PWD method to the curved line shown in Fig. 2. (Note that the absolute values of all matrix entries are 1.) Note the difference with respect to Fig. 5 in terms of sparsity.

As a proof of concept, we also test the developed formulations for 2D curved surfaces. We assume that both the 3D field and the curved surface are periodic along the x and y axes with the same period. The periodicity assumption is due to computational reasons. Note that with a proper indexing over the elements of the samples of the given field on the curved surface, we end up with a 1D array U . Using the same indexing, we obtain 1D arrays of field pattern samples that the expansion functions produce on the curved surface. These arrays are actually columns of the system matrix G . Thus, they form the G matrix. Therefore, the problem formulation given by Eq. (20) is also valid for curved surfaces.

We use a periodic curved surface, which we obtain by low-pass filtering a periodic 2D discrete function whose spectral coefficients are randomly chosen from the uniform distribution on the open interval (0,1) and a synthetically generated periodic object in 3D simulations. One period of the curved surface and object are shown in Figs. 7 and 8, respectively.

For the curved surface given by Fig. 7, the period along both the x and y axes is taken as 54.5λ . We discretize the curved surface by using a uniform transversal sampling grid where the sampling step is taken as 0.5λ along both the x and y axes. One period of the resulting 2D discrete signal is of size 108×108 . The total number of shift positions and the total number of modulation frequencies of the Gaussian window

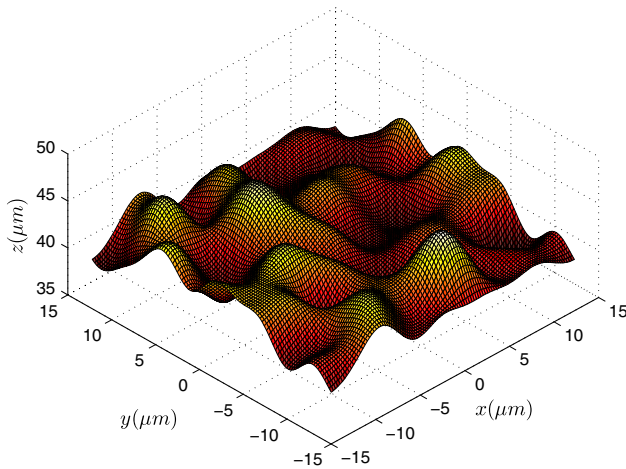


Fig. 7. (Color online) One period of the 2D periodic curved surface.

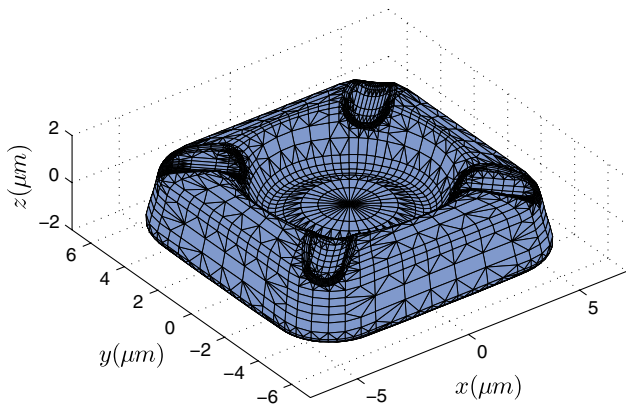


Fig. 8. (Color online) One period of the 2D periodic object. (Gilles Tran © 2007 www.oyonale.com, used under the Creative Commons Attribution license.)

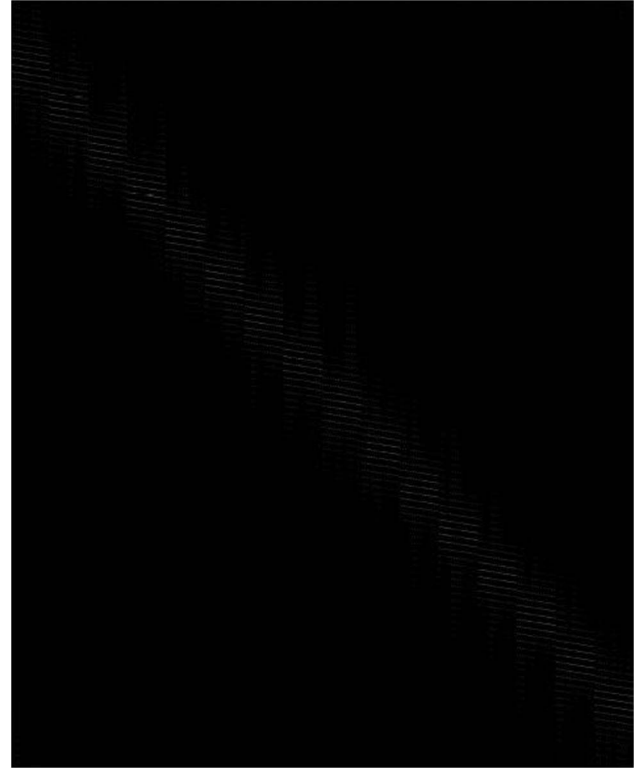


Fig. 9. Absolute value of the system matrix G , which has size 11664×9472 and is formed by the application of the proposed Gaussian beam decomposition method for the curved surface given in Fig. 7.

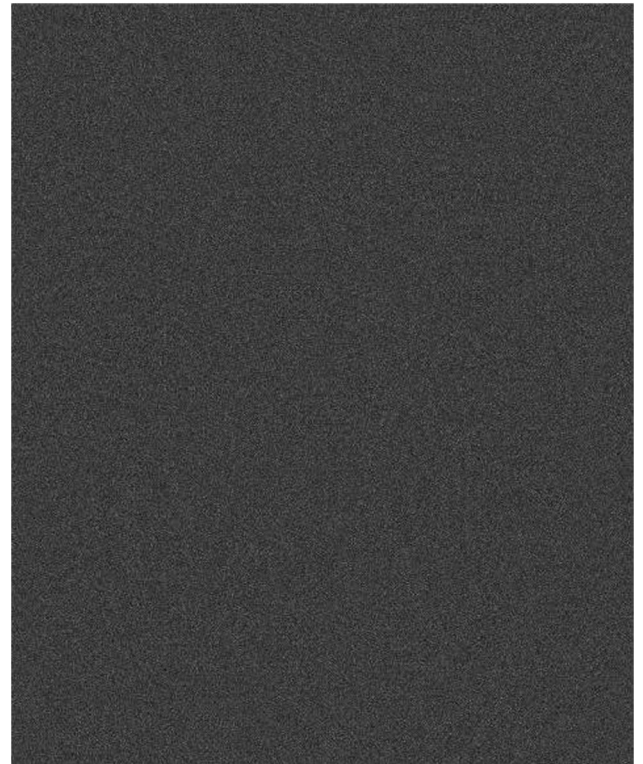


Fig. 10. Real part of the system matrix G , which has size 11664×9475 and is formed by the application of the PWD method for the curved surface given in Fig. 7. (The absolute values of all matrix entries are 1.) Note the random noiselike appearance, which is a consequence of the structure of the surface given in Fig. 7.

functions along the x and y axes are chosen to be the same. The total number of plane waves and Gaussian beams used in the decompositions are 9475 and 9472, respectively. Thus, noting that the total number of samples on one period of the curved surface is 11664, the system matrices are overdetermined.

The period along both the x and y axes is taken as 30λ for the object given by Fig. 8. The object is composed of 3973 vertices. In this simulation, the total number of plane waves and Gaussian beams used in the decompositions are 2819 and 2809, respectively. Therefore, the system matrices for both the Gaussian beam and the PWD methods are overdetermined.

By randomly choosing the coefficients of the periodic Gaussian beams [from the uniform distribution on the open interval $(0,1)$], we define the 3D field and form the corresponding system of linear equations. We repeat the proposed procedure several times for different randomly chosen 3D field examples. Examples of the system matrices related to the proposed Gaussian beam decomposition and PWD methods are shown in Figs. (9) and (10) for the curved surface given by Fig. 7. The system matrices for the object given by Eq. 8 are shown in Figs. 11 and 12. The sparsity of the system matrices corresponding to the Gaussian beam decomposition method are clearly seen in Figs. 9 and 11.

We solve the linear system of equations related to the proposed Gaussian beam decomposition and PWD methods by

applying the usual QR decomposition and multifrontal sparse QR factorization to the system matrices. For the curved surface given by Fig. 7, we report that the computational gain due to the usage of sparse QR factorization (with respect to the usual QR decomposition) is about 2 for the PWD method. However, this ratio is about 20 for the proposed method. When the sparse QR factorization is used, the time necessary to solve the linear system of equations related to the PWD method is about 10 times more than the time necessary to solve the linear system of equations related to the proposed Gaussian beam decomposition method.

For the object given by Fig. 8, if the sparse QR factorization is applied to the system matrix corresponding to the PWD method, the computation time is reduced by a factor of about 2 with respect to the usual QR decomposition. However, this reduction is by about a factor of 10 for the proposed method. When the sparse QR factorization is used, the time necessary to solve the linear system of equations corresponding to the PWD method is about 5 times more than the time necessary to solve the linear system of equations corresponding to the proposed Gaussian beam decomposition method. Thus, these simulation results show that the sparsity of the system matrices related to the proposed method are taken advantage of even for the small-scale examples that we demonstrate above.

The sparsity of the system matrix G is also beneficial in terms of memory. The sparsity of G corresponding to the proposed method relieves the memory requirement related to the

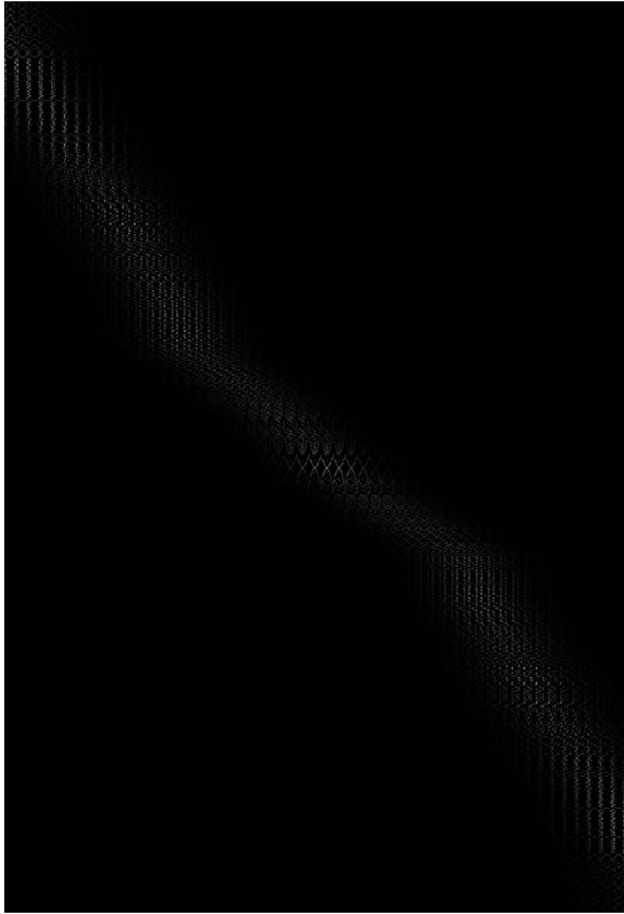


Fig. 11. Absolute value of the system matrix G , which has size 3973×2809 and is formed by the application of the proposed Gaussian beam decomposition method for the object given in Fig. 8.

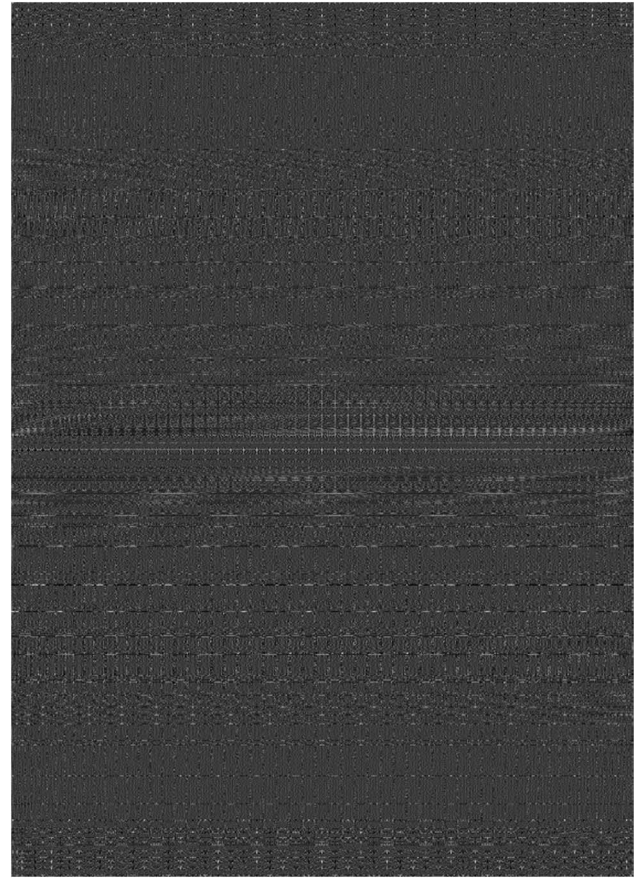


Fig. 12. Real part of the system matrix G , which has size 3973×2819 and is formed by the application of the PWD method for the object given in Fig. 8. (The absolute values of all matrix entries are 1.)

storage of the system matrix during the solution of the linear system of equations. In the proposed method, before applying the sparse QR factorization to the system matrix, thresholding is used, and then the system matrix is stored using the sparse storage scheme provided by MATLAB [20]. Once we have the matrix stored using the sparse storage scheme, we do not need the original matrix. Therefore, noting that the memory size of the original matrix is considerably higher than the memory size of the matrix stored using the sparse storage scheme, a memory gain is achieved. Even for the small-scale examples that we explain above, the memory requirement of the system matrix corresponding to the proposed Gaussian beam decomposition method can be reduced by a considerable amount. The reductions are by factors of 48 and 12 for the curved surface and the object given by Figs. 7 and 8, respectively.

5. CONCLUSIONS

A local Gaussian beam decomposition method is presented for the calculation of the exact scalar diffraction field from curved surfaces. The problem is formulated as a system of linear equations where the columns of the system matrix are the field patterns that the Gaussian beams produce on the curved surface. The proposed method provides a considerable amount of reduction in the computational complexity with respect to the method introduced in [9], which uses plane waves as the expansion functions. The reduction is a consequence of the sparsity of the system matrices related to the proposed method, and it becomes greater as the size of the curved surface gets larger.

In order to show the efficiency of the proposed method, the systems of linear equations related to the proposed Gaussian beam decomposition and the PWD [9] are solved by using the usual QR factorization method [22] and the multifrontal sparse QR factorization method [18]. For the curved line used in the simulations (see Fig. 2), these methods are applied several times to the system matrices having various sizes. On the average, when the sparse QR factorization is applied, the time necessary to solve the linear system of equations related to the PWD method is about 10 times more than the time necessary to solve the linear system of equations related to the proposed method for the system matrices having size 1200×1000 (see Fig. 3). This ratio becomes about 25 when the size of the system matrices is increased to 2400×2000 . Thus, the benefit of the sparsity in terms of computational complexity becomes more significant as the size of the curved surface increases. For the PWD method, the computational gain of the sparse solver with respect to the usual method using QR decomposition is about 2, when the size of the system matrix is 2400×2000 (see Fig. 3). However, the gain is about 60 for the proposed method. Thus, the sparsity of the system matrix related to the proposed method can be taken advantage of by a sparse solver.

The benefits of the proposed method are also observed in 3D simulations. For the curved surface shown in Fig. 7, the usual and sparse QR decomposition methods are applied several times for different 3D fields to the system matrices shown in Figs. 9 and 10. Even for this small-size curved surface, which is represented by 108×108 samples, the computational gain of the proposed method with respect to the PWD method is reported to be about 10 times when the sparse

solver [18] is used. Similar simulations are carried out for the object shown in Fig. 8 that is composed of 3973 vertices. When the sparse solver [18] is used for the system matrices shown in Figs. 11 and 12, the computational gain of the proposed method with respect to the PWD method is reported to be about 5 times.

Another benefit of the sparsity is the reduction in the memory requirement related to the storage of the system matrix during the solution of the linear system of equations. By using the sparse storage scheme provided by MATLAB [20], the reduction in the storage space that is required to store the sparse system matrices shown in Figs. 9 and 11 are reported to be by factors of about 48 and 12, respectively. Noting that these memory gains are achieved for small-scale examples; we expect the related memory gain to be much more for large sizes of curved surfaces.

The multifrontal sparse QR decomposition method [18] performs the matrix factorization by applying a sequence of operations to the smaller submatrices, which are called frontal matrices. Therefore, such an approach provides even more improvements in the computational complexity, if the parallelism among the frontal matrices is exploited by parallel processing.

For the problem of diffraction field calculation from curved surfaces, the approach given in [8] also suggests a Gaussian beam decomposition method and aims mainly to obtain a more accurate solution than the point-source models [1–7]. Although this approach provides quite accurate solutions, it imposes some restrictions on the curved surface, such as smoothness. On the other hand, the method that we propose in this paper does not impose any restriction on the curved surface and still provides exact scalar field solutions.

ACKNOWLEDGMENTS

Erdem Şahin acknowledges the partial support of TÜBİTAK for this work in the form of a scholarship.

REFERENCES

1. J. P. Waters, "Holographic image synthesis utilizing theoretical methods," *Appl. Phys. Lett.* **9**, 405–407 (1966).
2. T. Yatagai, "Stereoscopic approach to 3-d display using computer-generated holograms," *Appl. Opt.* **15**, 2722–2729 (1976).
3. K. Matsushima and M. Takai, "Recurrence formulas for fast creation of synthetic three-dimensional holograms," *Appl. Opt.* **39**, 6587–6594 (2000).
4. M. Lucente, "Optimization of hologram computation for real-time display," *Proc. SPIE* **1667**, 32–43 (1992).
5. K. Matsushima, "Computer-generated holograms for three-dimensional surface objects with shade and texture," *Appl. Opt.* **44**, 4607–4614 (2005).
6. M. Janda, I. Hanák, and L. Onural, "Hologram synthesis for photorealistic reconstruction," *J. Opt. Soc. Am. A* **25**, 3083–3096 (2008).
7. L. Ahrenberg, "Methods for transform, analysis and rendering of complete light representations," Ph.D. thesis (Max-Planck-Institut für Informatik, 2010).
8. E. Şahin and L. Onural, "Scalar diffraction field calculation from curved surfaces via Gaussian beam decomposition," *J. Opt. Soc. Am. A* **29**, 1459–1469 (2012).
9. G. B. Esmer, "Calculation of scalar optical diffraction field from its distributed samples over the space," Ph.D. thesis (Bilkent University, 2010).
10. G. B. Esmer, L. Onural, and H. M. Ozaktas, "Exact diffraction calculation from fields specified over arbitrary curved surfaces," *Opt. Commun.* **284**, 5537–5548 (2011).

11. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, 1996).
12. L. Onural, "Exact solution for scalar diffraction between tilted and translated planes using impulse functions over a surface," *J. Opt. Soc. Am. A* **28**, 290–295 (2011).
13. P. Flandrin, *Time-Frequency/Time-Scale Analysis* (Academic, 1999).
14. D. Gabor, "Theory of communication," *J. Inst. Electr. Eng.* **93**, 429–457 (1946).
15. M. J. Bastiaans, "Gabor's signal expansion and the Zak transform," *Appl. Opt.* **33**, 5241–5255 (1994).
16. M. J. Bastiaans, "Oversampling in Gabor's signal expansion by an integer factor," in *International Symposium on Time-Frequency and Time-Scale Analysis* (IEEE, 1994), pp. 280–283.
17. A. J. E. M. Janssen, "Gabor representation of generalized functions," *J. Math. Anal. Appl.* **83**, 377–394 (1981).
18. T. A. Davis, "Algorithm 915, SuiteSparseQR: multifrontal multithreaded rank-revealing sparse QR factorization," *ACM Trans. Math. Softw.* **38**, 1–22 (2011).
19. I. S. Duff, "A survey of sparse matrix research," *Proc. IEEE* **65**, 500–535 (1977).
20. J. R. Gilbert, C. Moler, and R. Schreiber, "Sparse matrices in MATLAB: design and implementation," *SIAM J. Matrix Anal. Appl.* **13**, 333–356 (1992).
21. E. Ulusoy, L. Onural, and H. M. Ozaktas, "Synthesis of three-dimensional light fields with binary spatial light modulators," *J. Opt. Soc. Am. A* **28**, 1211–1223 (2011).
22. G. W. Stewart, *Matrix Algorithms* (Society for Industrial and Applied Mathematics, 1998).